Geomagnetism in the satellite era

Citation for published version:
Whaler, K 2007, 'Geomagnetism in the satellite era', *Astronomy & Geophysics*, vol. 48, no. 2, pp. 23-29.
https://doi.org/10.1111/j.1468-4004.2007.48223.x

Digital Object Identifier (DOI):
10.1111/j.1468-4004.2007.48223.x

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Astronomy & Geophysics

Geomagnetism in the satellite era

Kathryn A Whaler considers the opportunities and challenges arising from the greatly enhanced geomagnetism data sets now available, in her 2006 Presidential Address.

Abstract

Following 20 years without satellite magnetic coverage, the first five years of the International Decade for Geopotential Field Research have provided the geomagnetic community with a wealth of high-quality data from several near-Earth satellites: Ørsted, SAC-C and CHAMP. Combined with ground-based and aeromagnetic data, this has opened numerous opportunities for studies ranging from core flow, mantle electrical conductivity, lithospheric composition and ocean circulation to the dynamics of ionospheric and magnetospheric currents using one or more satellites. Here, I review our current state of knowledge, and discuss the challenges to maximizing the utility of the satellite data.

We are in the International Decade for Geopotential Field Research, a worldwide effort to promote and coordinate continuous monitoring of the geopotential (magnetic and gravity) fields in the near-Earth environment. Already satellites have provided valuable new data that, in turn, are opening up new fields of research. In this review I will concentrate on the field sources internal to the Earth's surface, and what we can learn from them about the Earth's interior. I also include a comparison between the magnetic fields of Earth and Mars.

Data and modelling

Through extensive "data-mining" exercises, we now have observational magnetic field data and models from the 15th century onwards (e.g. Jackson *et al.* 2000 and described in Jackson *et al.* 1997). The early measurements, made for the purposes of navigation, were of declination (the angle between geographic and magnetic north), and later also of inclination (the angle a freely suspended magnet makes with the horizontal), extracted from ships' logs. In the mid-19th century, Gauss invented a method to measure the field intensity (described in this issue by Jackson, 2007), heralding the beginning of full vector definition of the field. He also established the first permanent magnetic observatories, a network of which now covers the globe (albeit with rather patchy distribution – poor over the oceans and in the southern hemisphere). These are supplemented by magnetic repeat stations, occupied periodically to improve coverage, and magnetic satellites.

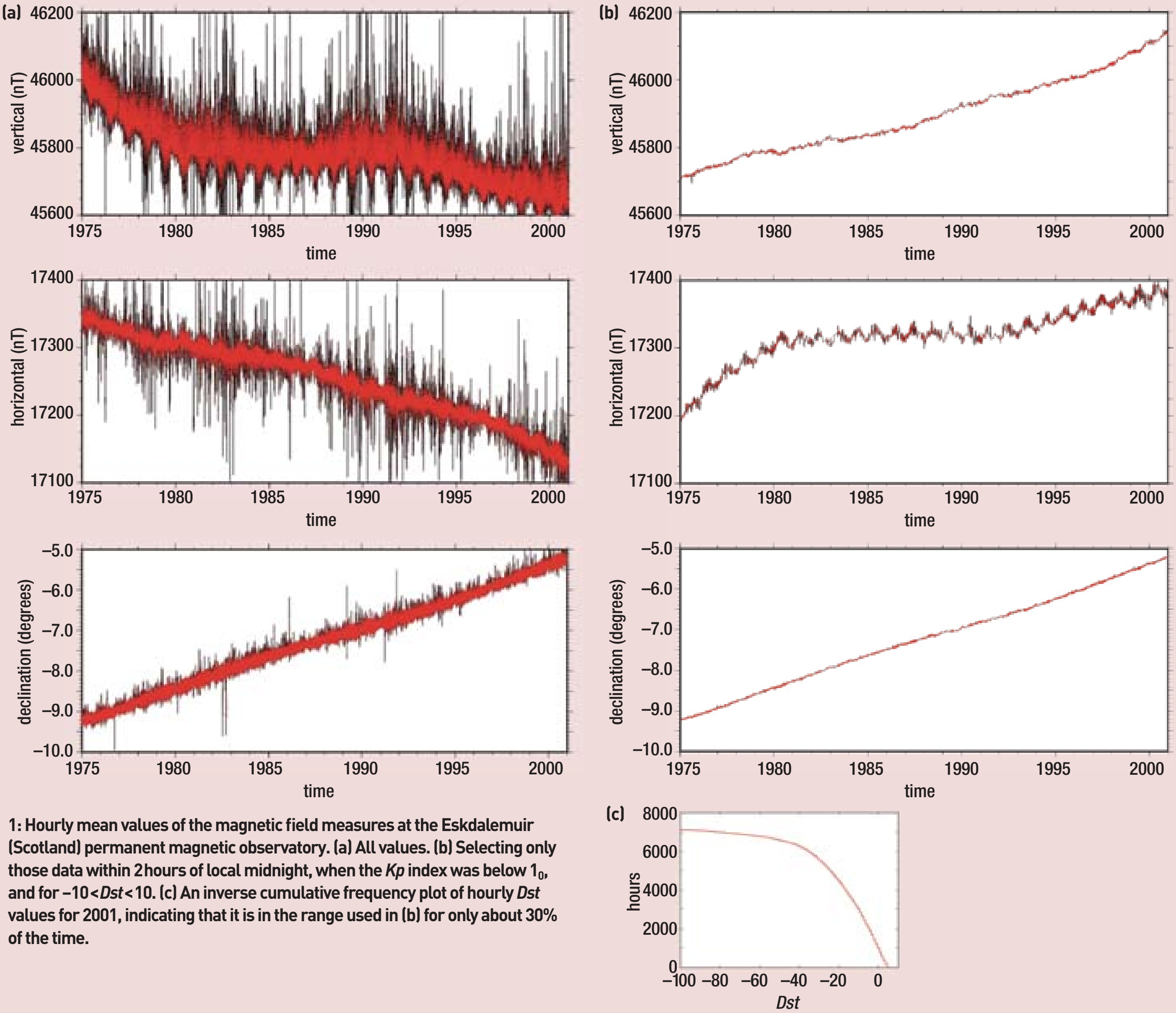
Modern satellites record three orthogonal vector field components, and a scalar intensity instrument for calibration and for measuring the field near the poles where vector data are less useful, whereas early missions deployed just a scalar magnetometer. The first vector satellite was MAGSAT, which operated for just seven months in 1979–80 in a Sun-synchronous orbit, measuring at local dawn and dusk. Currently, data are being recorded by the Ørsted, CHAMP and SAC-C satellites, which have been measuring continuously since Ørsted was launched in 1999. Their orbits cover all local times, and have varied in altitude between around 350 km (the current altitude of CHAMP) and almost 900 km (the initial apogee of Ørsted). Occasionally, aeromagnetic data, collected primarily for minerals and hydrocarbons exploration, are also used.

For older data, the uncertainties are sufficiently large, and the data sufficiently sparse – both spatially and temporally – that only the evolution of the field generated by self-sustaining dynamo action in the liquid iron core can be determined. This is known as the main field. There are two basic approaches to modelling the main field from data sets including satellite measurements, when fields arising from sources external to the Earth – the magnetosphere and ionosphere – must be taken into account. The first is to remove estimates of other field sources (see next section), or filter or average the data to minimize their influence, prior to modelling; the second is the so-called "comprehensive modelling" approach (e.g. Sabaka *et al.* 2002), where all sources are parameterized and solved for simultaneously. In both cases, it is common to select for modelling those data least affected by field sources external to the Earth (see next section). Models are expressed as spherical harmonic coefficient values.

Early modelling produced a series of "snapshots" of the field at different epochs, whereas computing power now enables time-dependent models of the field to be produced, capturing the main field and its temporal variation in a single set of parameters (e.g. Jackson *et al.* 2000). These models are designed to produce our best estimates of the field at the top of the source region, the core–mantle boundary (CMB), below which the expression of the field as the gradient of a scalar potential ceases to be a good approximation. Features can most easily be seen by viewing the models as "movies" showing the time evolution of the radial field component (e.g. http://earth.leeds.ac.uk/~earccf/animations/BzCS.gif); its sign indicates whether flux is entering or leaving the core, and its strength, the density of field line footprints on the CMB. To stabilize the downward continuation process (from at/near the Earth's surface where the data are collected to the CMB where the field maps are produced), a regularized least-squares approach to modelling is employed, where the objective function to be minimized includes a measure of spatial complexity as well as the sum of squares of the data residuals; rapid temporal variations are also mitigated against by a temporal complexity term in the objective function.

Sources of the field

The geomagnetic field measured at or near the Earth's surface comprises multiple, interacting sources generated by various mechanisms. They give rise to a field varying at an enormous variety of spatial and temporal scales, from the reversal of the dipole (bar magnet-like) component of the field originating in the core every half a million years or so on average, to micropulsations in the magnetosphere with periods of less than a second. No other measurable physical quantity can sense so many diverse regions of the Earth.



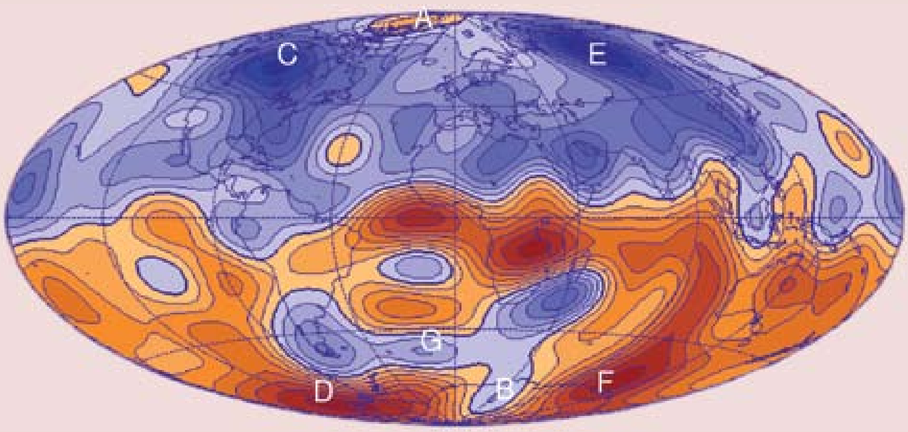
1: Hourly mean values of the magnetic field measures at the Eskdalemuir (Scotland) permanent magnetic observatory. (a) All values. (b) Selecting only those data within 2 hours of local midnight, when the *Kp* index was below 1_0, and for $-10 < Dst < 10$. (c) An inverse cumulative frequency plot of hourly *Dst* values for 2001, indicating that it is in the range used in (b) for only about 30% of the time.

The largest contributor, in terms of strength, is the main field, generated by self-sustaining dynamo action in the liquid iron alloy outer core of the Earth. At the Earth's surface, it has an amplitude of several tens of thousands of nanoTesla (nT), and is about 90% explained by a dipole at the Earth's centre, tilted about 11° from the rotation axis. The main field changes slowly with time – for example, the declination in the UK is currently about 3°W, changing by about 0.1°E per year. This is known as the secular variation. Most secular variation information is obtained from permanent magnetic observatories and repeat stations, but differencing main field models obtained from Ørsted and MAGSAT has provided a high-resolution model of average secular variation over the period 1980–2000 (Hulot *et al.* 2002). There is also strong evidence from observatory time series for occasional much more rapid variations, when the secular variation changes suddenly (i.e. on timescales of order months) from one approximately constant value to another, often with a sign change to give a V-shaped (or Λ-shaped) secular variation record. The secular acceleration consists of two offset straight line segments, a phenomenon known as a geomagnetic impulse or "jerk", by analogy with mechanics. In principle, jerks provide information on the mantle's electrical conductivity profile, as well as core dynamics, but there are complications and analyses to date have been relatively unsophisticated. As yet, there are only tentative observations of a jerk in satellite data.

The main field, now and in the past, is responsible for the lithospheric magnetic field, which reflects the magnetization (controlled primarily by mineralogy and temperature) of lithospheric rocks. Mapping the lithospheric field is used as a structural and tectonic tool, as well as in exploration for both minerals and hydrocarbons. The power spectrum of the magnetic field has a break in slope at around spherical harmonic degree 14, which is assumed to reflect the wavenumber beyond which the lithospheric field dominates over the main field. Thus lithospheric field maps are produced globally by including spherical harmonics starting from around degree 14, and locally by removing an estimate of the (long wavelength) main field over the survey region. The lithospheric field has an amplitude ranging from a few nT to typically several thousand nT over large mineral deposits, and varies on length scales characterized by the size of geological units. Magnetization has two components – remanent magnetization, parallel to the main magnetic field at the time the magnetization is locked in (e.g. as magma cools), and induced magnetization, parallel to the current main magnetic field. The relative importance of the two depends on the rock type and its history, with remanent magnetization typically more important in the oceans, and induced magnetization on the continents. The identification of magnetic stripes – alternately normal and reversed remanently magnetized ocean floor in bands parallel to and symmetric about the mid-ocean ridge – was of major importance in establishing the plate tectonics paradigm in the early 1960s. The alternating polarity of the stripes is evidence for magnetic polarity reversals. The remanent field is constant, but the induced field varies as the main field varies. Lithospheric field models based on CHAMP data now extend to spherical harmonic degree and order 100, providing unprecedented resolution (Maus *et al.* 2007).

External magnetic fields have traditionally been neglected, or treated only cursorily, by the solid-Earth geophysics community. They arise from the solar wind (high-speed charged particles emanating from the Sun) impinging on the internal magnetic field, which provides a shield out to many Earth radii protecting life on Earth from incoming radiation. The interaction between the solar wind and the internal field generates further magnetic fields in the magnetosphere and ionosphere according to Maxwell's laws. They tend to vary on much shorter temporal and spatial timescales than the main field, although there are several large-scale features such as ring currents and electrojets, and are much stronger on the day side of the Earth and near the poles. Solar activity is the main factor affecting external field strength. Those interested in the internal magnetic field tend to use data collected during "magnetically quiet" times and at night to minimize external field contamination. Magnetic activity is measured by a number of parameters, or indices, derived from magnetic observatory data which quantify certain aspects of the external field. On the other hand, electromagnetic induction, magnetospheric and ionospheric magnetic field, and what has become known as "space weather", studies depend primarily on data collected during "magnetically disturbed" times. Thus the internal and external magnetic field communities have had more-or-less mutually exclusive data interests – one scientist's noise has been another's signal.

As the external magnetic field changes, it induces electrical currents in the (weakly) conducting subsurface which, as they change, generate an internal induced magnetic field. The strength and geometry of the induced magnetic field depends on the electrical conductivity structure of the interior (as well as the strength of the external, inducing field). This is the basis of electromagnetic induction studies of electrical conductivity distribution, one aspect of which (magnetotellurics) was the subject of my Presidential Address last year (Whaler 2006). Satellite electromagnetic induction studies are in their infancy (e.g. Constable and Constable 2004, Balasis and Egbert 2006, Velimsky *et al.* 2006), but the expectation is that they will provide three-dimensional models of subsurface electrical conductivity, which in turn can be interpreted in terms of composition and temperature. However, a rather different induced signal, from motional induction by the lunar semidiurnal M_2 tide in the oceans, has now been identified in satellite data, despite only having an amplitude of about 1.5 nT (Tyler *et al.* 2003). It arises when electrically conducting seawater moves past magnetic field lines, so will be present even in a steady field.



2: Radial component of the magnetic field at the core surface (continents only for reference) in 1990 from model gufm (Jackson *et al.* 2000). Contour interval is 100 μT. Features labelled A–G are discussed in the text.

That we are now able to identify such tiny amplitude sources in the data is a testament to the quality of the acquisition, processing and modelling of the satellite data. A significant advance has been to treat properly the error budget when transforming the vector data from the satellite coordinate system to an Earth-centred one. Orthogonal fluxgate magnetometers are attached to the spacecraft boom, whose orientation is determined by one or more star cameras. This results in directional accuracy being better in some directions than others, and thus uncertainties on the different spacecraft field components are different; this anisotropy leads to a non-diagonal error covariance matrix when the data are rotated from the spacecraft reference frame to an Earth-centred one defining the usual north, east and vertically downwards field components. The improvement in main field models obtained from inverting data with a full error covariance matrix over those assuming isotropic errors is impressive (Holme 2000), and the additional computational burden is easily accommodated. If the data processing is treated in this more sophisticated fashion, it means in turn that a simplistic approach to data modelling – such as treating external field sources as zero mean, Gaussian distributed noise when modelling the main field – is no longer adequate. This has prompted the careful characterization of the fields from all possible sources, either to remove them prior to modelling, or to include them in a simultaneous inversion.

Separating the various sources, especially in satellite data, is by no means straightforward. For example, since the internal field induced by a changing external field has the same periodicity as the inducing field, the separation cannot be achieved on the basis of temporal variability. Mathematically, internal–external field separation is formally possible, but some sources (such as field aligned currents) are external to ground measurements, yet lie below the altitude at which satellites operate. In principle, data recorded at different satellite altitudes can help discriminate between sources. However, it is extremely hard to distinguish between rapid temporal and rapid spatial field variations from a single, fast-moving satellite. Further complications arise when we attempt simple parameterizations of the various field sources, based upon what we know or expect of their geometry. These parameterizations are only simple in an appropriate coordinate system, which may differ for the different sources. Thus we have field sources that can be described straightforwardly as the gradient of a potential containing only a few important terms in one coordinate system, but are non-potential in the coordinate system that naturally describes another source. Also, the relationship between the various co-ordinate systems is generally time-dependent.

The amplitude of the magnetic field is not its only important attribute. For example, the strength of geomagnetically induced currents (GIC) in power transmission lines depends primarily on the field's rate of change with time. If large enough, these GIC can trip switches designed to protect the transmission system and cause the supply to fail, as most famously happened in Quebec, Canada, during a major magnetic storm in 1989. GIC are influenced by the local electrical conductivity structure, which in the UK depends on proximity to the coast (and bathymetry) as well as the geology (McKay and Whaler 2006).

The selection criteria we use to choose data from magnetically quiet periods for main field modelling result in over 90% of the available data being discarded, to reduce the noise in the data (see figure 1). For example, we tend to restrict data to those obtained close to midnight local time, and when the *Dst* and *Kp* indices (also, their time rate of change) are below certain values (see figure 1). Even extending modelling to include data obtained during moderately disturbed times would improve the applicability of the models, and allow us to merge different types of study and data, particularly through taking advantage of the expertise in and data available for studying the external field (which is aimed primarily at the disturbed field). This could be important for a better understanding of phenomena such as GIC, and also areas of the globe where external fields are more important, such as over the poles and the "South Atlantic Anomaly" (described below).

Deep Earth inferences

The structure of the CMB field and its secular variation include static features, drifting features, regions of lower than average or expected field strength or secular variation, and suggestions of wave-like features. Prominent static features are low intensity over the poles (features A and B in figure 2), and two pairs of oppositely signed high flux lobes separated by about 120° in longitude (features C–F in figure 2). The low intensity over the poles is contrary to expectations, given that a slightly tilted dipole field that reaches its maximum near the poles fits over 90% of the field at the Earth's surface.

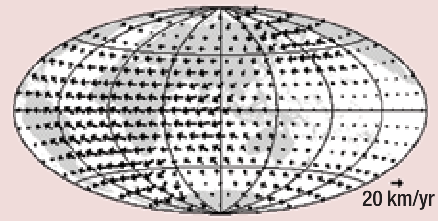
On average, the field drifts west at a rate of about 0.2°/year; westward drift was first observed by Halley, who made a series of measurements across the Atlantic Ocean. However, had Halley lived instead in the Pacific hemisphere, he would have observed no, or possibly a weak eastward, drift. Secular variation in general is much lower over the Pacific hemisphere. Both mid-latitude (Bloxham *et al.* 1989) and equatorial (Finlay and Jackson 2003) westward-propagating waves have been proposed, concentrated in the Atlantic hemisphere. Hide (1966) first suggested wave propagation as an explanation for westward drift, as an alternative to Bullard *et al.*'s (1950) proposal that it arises from differential rotation in the outer core.

Magnetic field evolution can be used to probe core dynamics by noting that, on timescales of decades, we expect advection to dominate over diffusion, owing to the high electrical conductivity of the outer core liquid iron alloy. In the limit of perfect conductivity, this results in the magnetic field being frozen into the fluid. Thus we can use field lines as tracers of the flow. In this "frozen flux" approximation, only the radial field component is guaranteed continuous across the conductivity jump at the CMB.

In a rapidly rotating annulus (such as the Earth's outer core), thermal convection takes the form of rolls parallel to the rotation axis and tangential to the inner core boundary at the equator (Busse 1970). Thus they terminate where the projection of the inner core boundary along the rotation axis meets the core–mantle boundary, at ~70° North and South, defining a region within known as the tangent cylinder. Adjacent "Busse rolls" have fluid alternatively spiralling into or spun out of their ends. The latitudes of the stationary high flux lobes (c–f of figure 2) coincide with the edge of the tangent cylinder. Thus they have been proposed to represent the ends of convection rolls into which fluid is spiralling, dragging field lines with them and hence concentrating the flux. To have only two such rolls implies a far lower dominant wavenumber than that found in purely thermal convection. In theory, they should also be equally spaced in longitude; this suggests either that there is a third roll missing, or that the pattern is wavenumber 2 with the rolls having drifted closer together. Another aspect in which the roll explanation of the high flux lobes differs from theory is that the whole pattern should drift azimuthally, whereas the lobes have been stationary for the duration of the historical record. A possible locking mechanism is thermal control by the mantle (e.g. Zhang and Gubbins 1996) – mantle convection is expected to result in lateral variations in CMB heat flux, which may influence where the ends of rolls form. As mantle convection operates on a vastly slower timescale than core convection, the pattern of CMB heat flux, and hence the roll ends, will be effectively fixed when compared to the predicted drift rate.

Further evidence for mantle control on core processes comes from the palaeomagnetic reversal record. From an oriented rock sample, it is possible to calculate the position of the magnetic pole when the magnetic field was recorded by the rock, under the assumption that it was a dipole aligned with the Earth's centre (e.g. Merrill *et al.* 1996). This location is known as a "virtual geomagnetic pole", or VGP. In general, VGPs are close to either the north or south geographic pole, depending on whether the sample was from a period when the field was in normal or reversed polarity. During a reversal, the VGPs trace a path between the geographic poles. Some studies have found that these VGP paths are concentrated near the longitudes of the high flux patches (Laj *et al.* 1991), rather than being randomly distributed. Additionally, dynamo simulations with heterogeneous heat flux outer boundary conditions can give very different, in some cases more Earth-like, reversal records from those with homogeneous heat flux (Glatzmaier *et al.* 1999). However, the field reverses on average every half a million years or so, and the duration of a reversal is only ten thousand years or so. Thus our chances of sampling it during a reversal are poor, and simulations suggest that the "preferred VGP paths" might result from palaeomagnetic sampling position bias (e.g. Love 2000).

As already mentioned, the CMB field has considerably more structure than the almost dipolar field at the Earth's surface. In particular, there are areas where the radial component has the "wrong" sign (reversed flux in comparison to that of a dipole), notably in the patch beneath South American, the southern Atlantic Ocean, and South America (feature G on figure 2). At the Earth's surface, this patch contributes



3: Flow at the core–mantle boundary (continents only for reference) that explains the majority of the observed secular variation in the frozen-flux approximation.

to one of the main departures from the tilted dipole model, a region of considerably depressed (around 35% lower than expected) field intensity – the so-called South Atlantic Anomaly (SAA). The SAA is growing (i.e. the field strength is decreasing) at a significant rate, and is the major contributor to the decay of the dipole, whose strength has reduced by 10% over the last 150 years. Although dynamo field generation is highly nonlinear, so we do not necessarily expect decay to continue at its present rate (or even to continue at all), geomagnetists are beginning to wonder whether this heralds a field polarity reversal; given that the last reversal was ~0.75 million years ago, the Earth is statistically "overdue" for one. If the flow in the core driving the dynamo were turned off, the dipole decay rate (by diffusion) would be 10 times slower than the currently observed rate, so it is an active process, and is commensurate with reversal rates deduced from the palaeomagnetic record. Whether or not the field polarity is about to reverse, the SAA depressed field strength is an environmental hazard, as this region is less well shielded from space radiation. Low-Earth-orbit satellite failures and anomalies are concentrated in this region, so some satellites lose, do not collect (to protect detectors) or do not use (because of possible anomalies) data during their passage across the SAA. We know also that increased particle flux increases the air drag on satellites, distorting their paths and increasing fuel usage. In principle, the area could become dangerous to aircraft and their occupants because of the radiation dosage. Thus there are societal reasons for trying to understand the processes causing the SAA and predicting how deep it will become, and when it will stop growing.

CMB velocity maps obtained using field lines as flow tracers in the frozen-flux approximation were first calculated several decades ago (e.g. Kahle *et al.* 1967). Subsequently, Roberts and Scott (1965) and Backus (1968) pointed out the serious ambiguity involved. From a mathematical viewpoint, this arises because there are two unknowns, the tangential velocity components (the radial component vanishes because the CMB is a material boundary), to be derived from only one equation, the radial component of the induction equation. Physically, the flow non-uniqueness arises because we cannot label individual field lines that act as flow tracers. This ambiguity is in addition to that arising because we have only a finite quantity of imperfect data from which to calculate our model. Several flow non-uniqueness-reducing assumptions have been proposed, which can be tested for consistency against the data. These include that the flow is steady (Voohies and Backus 1985), tangentially geostrophic (Hills 1979, Le Mouël 1984), purely toroidal (Whaler 1980) or has a particular helicity (Amit and Olson 2004). Weak poloidal flow might indicate a thermally or chemically stratified layer adjacent to the CMB as the Earth cools, e.g. from the light material rejected as a denser fraction preferentially freezes to form the inner core. Helical flow is known to be an important ingredient of dynamo action. Tangentially geostrophic flows can be used to probe torsional oscillations in the core (e.g. Zatman and Bloxham 1997), which are the first response of the core to an applied torque.

Most calculations are based on spectral methods involving a decomposition of the flow into its toroidal and poloidal components expressed through scalars satisfying Laplace's equation, but a few use local methods. In fact, the precise nature of the non-uniqueness reducing assumption does not seem to be important, since most flows share broadly similar features: even if no additional constraints are imposed, flows tend to be predominantly toroidal and tangentially geostrophic (these components represent approximately 80% of the flow energy [Bloxham 1989]), and do not change rapidly with time. Steady flows are able to represent the gross features of the secular variation, but cannot reproduce fine-scale temporal variations; however, if the reference frame in which the steady flow is derived is allowed to drift with respect to the mantle, they can achieve a fit almost as good as a fully time-dependent flow, but expressed with only a fraction of the number of parameters (Holme and Whaler 2001). The practical uniqueness does not seem to depend on regularization (applied to spectral methods), because flows produced using local methods show broadly similar features. All non-uniqueness-reducing assumptions can only be approximate, and break down in different ways, but in general the data seem to be almost as consistent with them as they are with the underlying frozen-flux hypothesis.

An example velocity map is presented in figure 3. Flow speeds at the equator are consistent with the observed westward drift rate at the Earth's surface, and there is a well-defined band of westward-drifting flow near the equator over the Atlantic hemisphere dragging field features with it. Jets from the poles feed this band at around 90°E. Velocities tend to be lower beneath the Pacific Ocean, as expected because of the lower secular variation there. A large anti-clockwise gyre occurs beneath the south-west Indian Ocean. It is difficult to identify places at which flow is converging or diverging (poloidal flow) because the toroidal flow component is much stronger; poloidal flow is also much less consistent between models.

Hulot *et al.* (2002) inferred a finer scale tangentially geostrophic flow model from their satellite-based, high-resolution secular variation model which suggested a distinction between flow inside and equator-wards of the tangent cylinder. Within the tangent cylinder, the axisymmetric component of the flow is strongly westward, suggesting polar vortices, whereas it almost vanishes in the equatorial region. The non-axisymmetric component has small-scale prograde and retrograde vortices concentrated at the edge of the tangent cylinder reminiscent of those associated with Busse rolls produced by numerical geodynamo simulations; in contrast to the field models, which suggest just two pairs of rolls, the high-resolution flow model has many vortices, in common with dynamo models. Unfortunately, they are affected particularly severely by the inherent flow non-uniqueness, so their interpretation, and even their existence, is uncertain.

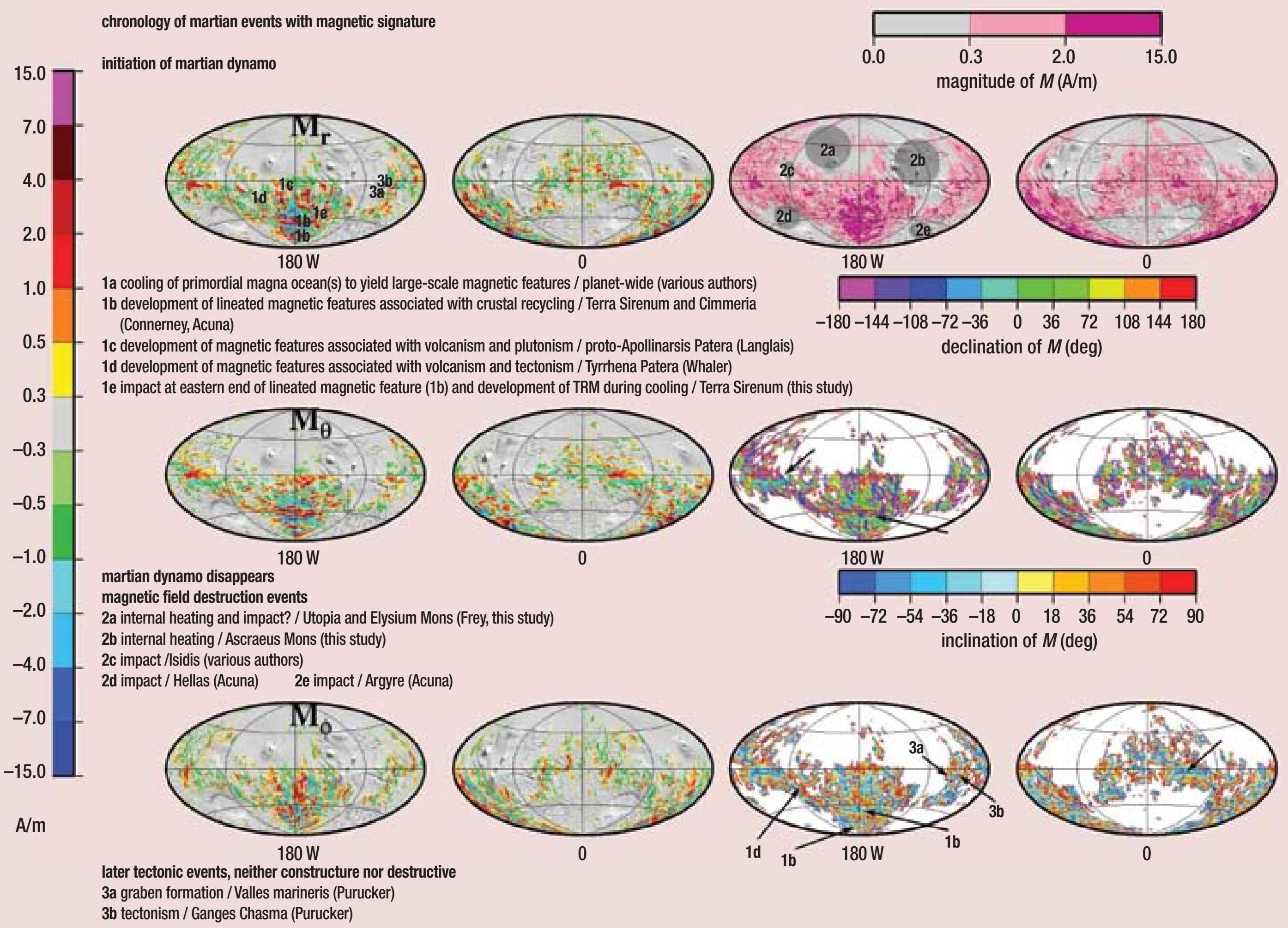
The inherent ambiguity in all flow modelling, and the need to impose the frozen-flux constraint, means that we cannot be sure to what extent the flows deduced represent the actual flow at the core surface, even if they provide a perfect fit to the secular variation. Studies aimed at recovering synthetic flows from the predicted field evolution in dynamo simulations have been equivocal (e.g. Rau *et al.* 2000). However, the decadal timescale changes in length of day, which we believe to arise from exchange of angular momentum between the core and mantle, are well represented by core flows deduced geomagnetically (Jault *et al.* 1988, Jackson *et al.* 1993, Holme and Whaler 2001). The length-of-day record is an independent data set, so our ability to match it with core flows gives them credence.

Our ability to match the decadal length-of-day record with core flows demonstrates kinematically the plausibility of angular momentum exchanges between the core and mantle, but does not address the dynamical problem of what the coupling mechanism between them is. The three most likely contenders are electromagnetic, gravitational and topographic core–mantle coupling. In electromagnetic coupling, we recognize that currents leak into the (weakly) conducting lower mantle, and the toroidal magnetic field is not confined to the core. Gravitational coupling arises from the attraction between density inhomogeneities in the mantle and core, most likely the inner core as the outer core should be well-mixed. Topography on the core–mantle boundary will provide a physical couple between them. Whereas it is hard to go beyond order-of-magnitude estimates for gravitational and topographic coupling, with many key quantities unknown, it is possible to demonstrate that electromagnetic coupling is viable (e.g. Holme 1998). The torque predicted by the length-of-day changes is of order 10^{17} Nm, and this requires a layer of conductance (conductivity-thickness product, measured in Siemens) 10^{8} S adjacent to the CMB. If we identify this layer with the D″ region, the extremely heterogeneous thermal and chemical boundary layer at the base of the convecting mantle, with a (highly variable) thickness up to a few hundred kilometres, then the implied electrical conductivity is a few hundred Siemens per metre, at least an order of magnitude higher than that inferred for lower mantle minerals (e.g. Xu *et al.* 2000). However, it has recently been shown that the silicate mineral perovskite (Pv), thought to be a major constituent of the lower mantle, transforms to a new phase known as post-perovskite (pPv) at temperatures and pressures inferred to exist in the D″ region. Because pPv can contain higher proportions of iron than Pv, its electrical conductivity may be higher (Hirose 2006).

Comparative planetology

The instrumentation, techniques and methods used by geophysicists (and geologists and geochemists) to study the Earth can now be applied to other solar system objects. A case in point is Mars – the orbiting satellite Mars Global Surveyor (MGS) collected data, including spectrometer, laser altimeter and magnetic, from the planet between 1997 and 2006, and surface rovers have been photographing, sampling and testing the rocks on the surface since 2004. MGS included an oriented vector magnetometer, which has provided three-component magnetic field data from altitudes between approximately 100 and 400 km. Previous fly-by missions had indicated that Mars no longer had a main field; this was confirmed by MGS. It is now widely accepted that the martian dynamo ceased about half a billion years after the planet's formation. MGS has been able to map the crustal magnetic field, representing remanent magnetization, in detail, and these maps have produced many surprises, and provided clues about Mars's history.

The topography of Mars is high and relatively low relief north of the so-called dichotomy, and lower lying and heavily cratered to the south of it. The topographic difference is also reflected in the magnetic field, which is weak to the north and much stronger to the south of the dichotomy, although in both hemispheres some large craters are non-magnetic. Non-magnetic craters can help determine when Mars' dynamo switched off, since they must post-date it: any pre-existing magnetization would be destroyed by the heat generated by the impact, but if the main field still existed, its signature at that time would be imprinted on the rocks as they cooled. One of the early spectacular discoveries of MGS was a series of "magnetic stripes" – linear (in an appropriate map projection) bands of positive and negative vertical magnetic



4: Martian magnetization superimposed on topography (from the MGS laser altimeter) shown in shaded relief. Left panels show the radial, north and east components (M_r, M_θ and M_ϕ respectively), according to the scale on the left. Right panels show amplitude and angles, with angles plotted only where the strength is large enough for them to be determined reliably. Features labelled are interpreted in a possible magnetic chronology. (Whaler and Purucker 2005)

field over the southern hemisphere (Connerney *et al.* 1999). They are somewhat reminiscent of terrestrial ocean magnetic stripes, and have led to speculation about plate tectonic-like motions on Mars. Moreover, if this is the correct interpretation, they indicate that the martian magnetic field reversed polarity during its short lifetime.

Another unexpected feature is that the martian crustal magnetic field is approximately an order of magnitude stronger than that of the Earth. This is not because it is closer to the sources, despite Mars being smaller – the attenuation depends on the ratio of the source to the observation radius, and for both the Earth and Mars, the core is roughly half the planet's radius. Models of the distribution of magnetization in the crust (e.g. Whaler and Purucker 2005) have features that could be interpreted structurally and tectonically. These include offsets that could result from movement on faults, and patterns of magnetization that would occur over a triple junction developing in a reversing magnetic field, and VGPs that indicate that rock units have either moved or rotated, or were not magnetized in a magnetic field that was predominantly dipolar and aligned with the rotation axis. Figure 4 plots a magnetization model, identifies some of the features referred to above, and gives a possible magnetic chronology. Given that the martian field strength is larger, it is not surprising that the magnetization values are higher than those typical (from direct measurement and inferred from modelling magnetic field measurements) of the Earth.

One of the intriguing questions to arise from the MGS measurements is why the crustal field strength of Mars is so much higher than that of Earth; another is why the two planets' magnetic histories are so different. Tidal data indicate that Mars still has a fluid core (Yoder *et al.* 2003), so the dynamo did not cease because its core froze. However, approximately 4 billion years ago, it ceased convecting, at least in a manner that generated a magnetic field. We would expect Mars to cool more quickly than the Earth because of its smaller size, but the magnetic field data indicate a very different thermal history from that of the Earth (see Stevenson 2001 for possible models). Perhaps the dynamo was more vigorous during its short lifetime, and used up its energy sources more quickly, generating a more powerful field while it did operate. However, we also expect a higher percentage of iron in the crust, which would contribute to the field being stronger, and it may have a different mineralogy.

Plans are underway to use ESA's ExoMars mission to deploy a magnetometer on the surface of Mars. Comparison between surface measurements on Earth extrapolated to satellite altitude and those actually measured by satellites indicate a mismatch – there is less power in the satellite measurements than the upward-continued surface data, indicating that the spectrum is not white. While a single surface magnetometer cannot establish whether the same is true of the martian magnetic field, unless it were also able to record during descent, it will give a direct measurement of the surface field strength, and start to give an indication of the form of the power spectrum. A permanent magnetometer on the surface would also establish whether the assumptions we have made about the contribution the external magnetic field has made to the measurements are warranted, as well as provide data on "space weather" on Mars.

The new satellite era

Our desire to understand the dynamo process responsible for generating our magnetic field (and the absence of a dynamo in other solar system objects we might expect to have a main field), and the need to continue to monitor the magnetic field from space because of the hazards associated with it (and the possible impact on climate), have meant that the geomagnetic community has successfully proposed a new mission, SWARM, to ESA, for launch in 2010. SWARM will be a constellation of three satellites in low-altitude, near-polar orbits, two at lower altitude flying closely side-by-side, the third higher, initially roughly antipodal to the other two. It is intended that gradients of the field will be estimated from the lower satellites, improving its resolution. The higher and lower orbit satellites will drift at different rates owing to their different inclinations.

The original SWARM configuration involved four satellites, with the two lower altitude ones flying one closely behind the other on the same orbit. However, as part of the preparation of the case for the mission, an "end-to-end simulator" study was undertaken (see the series of articles in *Earth, Planets and Space* 2006 **58** 349–496), in which satellite data containing contributions from multiple sources were synthesized, noise added, and attempts made to determine the original source signatures from the synthetic data using available modelling strategies. This study demonstrated that three satellites with the lower two flying side-by-side was at least as effective a configuration as that originally proposed – and considerably cheaper. It also indicated that the mission would be able to resolve the various field sources.

SWARM will provide the best ever survey of the geomagnetic field and its temporal evolution, and gain new insights into improving our knowledge of the Earth's interior and climate. The geomagnetic field models resulting from the SWARM mission will further our understanding of atmospheric processes related to climate and weather and will also have practical applications in many different areas, such as space weather and radiation hazards. GPS receivers, an accelerometer and an electric field instrument will provide supplementary information for studying the interaction of the magnetic field with other physical quantities describing the Earth system – for example, SWARM could provide independent data on ocean circulation.

Future developments

Current satellite missions and associated advances in analysis methods have provided significantly improved models of all sources contributing to the measured magnetic field, and stimulated research to understand the processes involved. Some signals, e.g. associated with ocean tides, have been isolated for the first time. Our aims include extending the resolution of the secular variation, which is not contaminated by the lithospheric field, out to higher spherical harmonic degree, to allow a better comparison with the output of numerical dynamo models. Developments currently underway include further applications of one-norm modelling, including in-core surface flow modelling, since data residuals seem to be better described by a Laplacian than a Gaussian distribution, continued exploration of maximum entropy modelling, and co-estimation of the Euler angles describing the transformation of the satellite reference frame to an Earth-centred reference frame along with the Earth-centred frame field components themselves (including error anisotropy). In addition, we are exploring whether scalar data from the POGO series of satellites in 1965–71 can be reprocessed taking advantage of recent advances to extend satellite-based time-dependent field models back in time. Besides contributing to our current knowledge, all these developments will maximize the value to be obtained from the forthcoming SWARM mission.

Conclusions

We are now closer than ever to understanding the mechanisms through which our planet's geomagnetic field is generated and maintained through self-sustaining dynamo action in the liquid outer core. Satellite data have been instrumental in providing the high-resolution images of the field at the core–mantle boundary, first as a snapshot and now with time dependence, necessary to do this. In addition, we have isolated and learned a great deal about other sources contributing to the measured magnetic field, which provide further information on the Earth's structure and constitution. Comparisons with satellite data from other solar system bodies, which can now be treated by methods developed for terrestrial data, are providing further insights. ●

Kathryn A Whaler is Professor of Geophysics at the Grant Institute of Earth Science at the University of Edinburgh.

Acknowledgments. Many people have stimulated and maintained my interest in various aspects of geomagnetism over the years, beginning with my PhD supervisor David Gubbins and subsequently including: the late Edward Benton; Jeremy Bloxham; past and present members of the geomagnetism group at GeoForschungsZentrum, Potsdam; Richard Holme and Andrew Jackson, and their post-doctoral researchers and PhD students; David Kerridge and the British Geological Survey geomagnetism group; the late Robert Langel, Michael Purucker and Terry Sabaka at NASA's Goddard Space Flight Center; past and current geomagnetists at the Institut de Physique du Globe, Paris, led until recently by Jean-Louis Le Mouël; Nils Olsen; Robert Parker and colleagues at the Institute of Geophysics and Planetary Physics, University of California at San Diego; and Tiku Ravat. This work is funded through a NERC Consortium grant to the universities of Edinburgh, Leeds and Liverpool, the British Geological Survey, and the Council for the Central Laboratory of the Research Councils; in addition, researchers from the University of Washington, Seattle, the Danish National Space Center, and ETH, Zürich, are overseas Project Partners.

References

Amit H and Olson P 2004 *PEPI* **147** 1–25.
Backus G E 1968 *Phil. Trans. R. Soc. Lond.* **A263** 239–266.
Balasis G and Egbert G D 2006 *Geophys. Res. Lett.* **33** L11311 doi:10.1029/2006GL025721.
Bloxham J *et al.* 1989 *Phil. Trans. R. Soc. Lond.* **A329** 415–502.
Bloxham J 1989 *Geophys. J. Int.* **99** 173–182.
Bullard E C *et al.* 1950 *Phil. Trans. R. Soc. Lond.* **A243** 67–92.
Busse F H 1970 *J. Fluid Mech.* **44** 441–460.
Connerney J E P *et al.* 1999 *Science* **284** 794–798.
Constable S and Constable C 2004 *Geochem. Geophys. Geosyst.* **5** Q01006 doi:10.1029/2003GC000634.
Finlay C C and Jackson A 2003 *Science* **300** 2084–2086.
Glatzmaier G A *et al.* 1999 *Nature* **401** 885–890.
Hide R 1966 *Phil. Trans. R. Soc. Lond.* **A259** 615–650.
Hills R G 1979 Convection in the Earth's mantle due to viscous shear at the core-mantle interface and due to large-scale buoyancy, PhD thesis, New Mexico State University, Las Cruces.
Hirose K 2006 *Rev. Geophys.* **44** RG3001 doi:10.1029/2005RG000186.
Holme R 1998 *GJI* **132** 167–180.
Holme R 2000 *Earth, Planets, Space* **52** 1187–1197.
Holme R and Whaler K A 2001 *GJI* **145** 560–569.
Hulot G *et al.* 2002 *Nature* **416** 620–623.
Jault D *et al.* 1988 *Nature* **333** 353–356.
Jackson A *et al.* 1993 in *Dynamics of the Earth's Deep Interior and Earth Rotation* eds Le Mouël J-L, Smylie D E and Herring T (AGU/IUGG, Washington, DC) 97–107.
Jackson A *et al.* 1997 *A&G* **38** 10–15.
Jackson A *et al.* 2000 *Phil. Trans. R. Soc. Lond. A* **358** 957–990.
Jackson A 2007 *A&G* **48** 2.9–2.13.
Kahle A B *et al.* 1967 *J. Geophys. Res.* **72** 1095–1108.
Laj C *et al.* 1991 *Nature* **351** 447–447.
Le Mouël J-L 1984 *Nature* **311** 734–735.
Love J J 2000 *GJI* **140** 211–221.
Maus S, Lühr H, Rother M, Hemant K, Balasis G, Ritter P and Stolle C 2007 Fifth generation lithospheric magnetic field model from CHAMP satellite measurements, *Geochem. Geophys. Geosyst.* submitted.
McKay A J and Whaler K A 2006 *GJI* **167** 613–625.
Merrill R T *et al.* 1996 *The Magnetic Field of the Earth: Paleomagnetism, the Core, and the Deep Mantle* (Academic Press, San Diego, California) 531.
Rau S *et al.* 2000 *GJI* **141** 485–497.
Roberts P H and Scott S 1965 *J. Geomag. Geoelectr.* **17** 137–151.
Sabaka T J *et al.* 2002 *GJI* **151** 32–68.
Stevenson 2001 *Nature* **412** 214–219.
Tyler *et al.* 2003 *Science* **299** 239–241.
Velimsky J *et al.* 2006 Electrical conductivity in the Earth's mantle inferred from CHAMP satellite measurements – I. Data processing and 1-D inversion *GJI* **166** 529–542.
Voorhies C V and Backus G E 1985 *Geophys. Astrophys. Fluid Dyn.* **32** 163–173.
Whaler K A 1980 *Nature* **287** 528–530.
Whaler K A 2006 *A&G* **47** 2.28–2.37.
Whaler K A and Purucker M E 2005 *J. Geophys. Res.* **110** E09001 doi:10.1029/2004JE002393.
Xu Y S *et al.* 2000 *J. Geophys. Res.* **105** 27865–27875.
Yoder C F *et al.* 2003 *Science* **300** 299–303.
Zatman S and Bloxham J 1997 T *Nature* **388** 760–763.
Zhang K, Gubbins D 1996 *Phys. Fluids* **8** 1141–1148.